# Everyday Grocery Sales Data Analysis

## 1. Introduction

The objective of this project is to analyze grocery sales data to understand patterns related to product attributes and outlet characteristics that influence total sales. In particular, the focus is on fat content preferences, product category performance, outlet location and size impact, and sales trends by outlet age. These insights can help improve marketing, inventory management, and strategic decisions.

## 2. Dataset Overview

The dataset contains detailed transactional sales information, including:

Product features like Item Fat Content, Item Type, Item Weight, and Rating,Outlet characteristics such as Outlet Location Tier, Size, Type, and Establishment Year

Sales data per item-outlet combination

This data allows multi-faceted analysis of sales drivers across customer preferences and store properties.

## 3. Methodology

Imported and cleaned the dataset using **Pandas**.

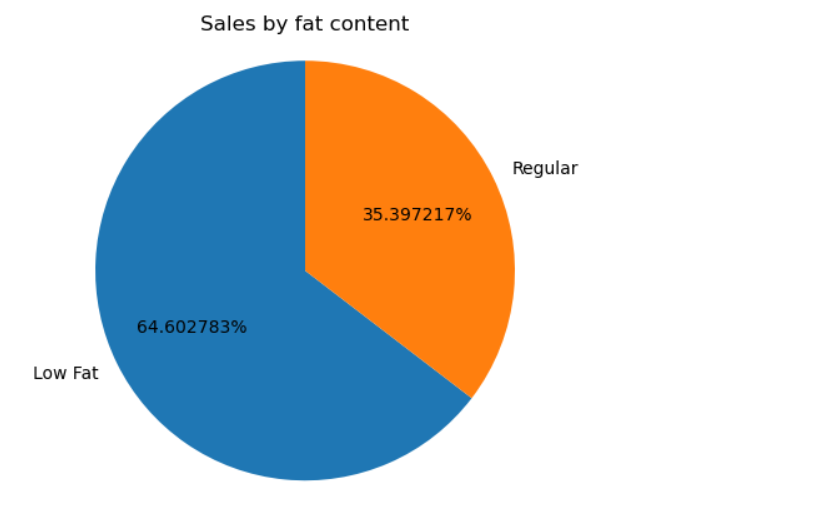
Performed data aggregation with grouping and summation by relevant columns.

Created visualizations using **Matplotlib** and **Seaborn** for better insight.

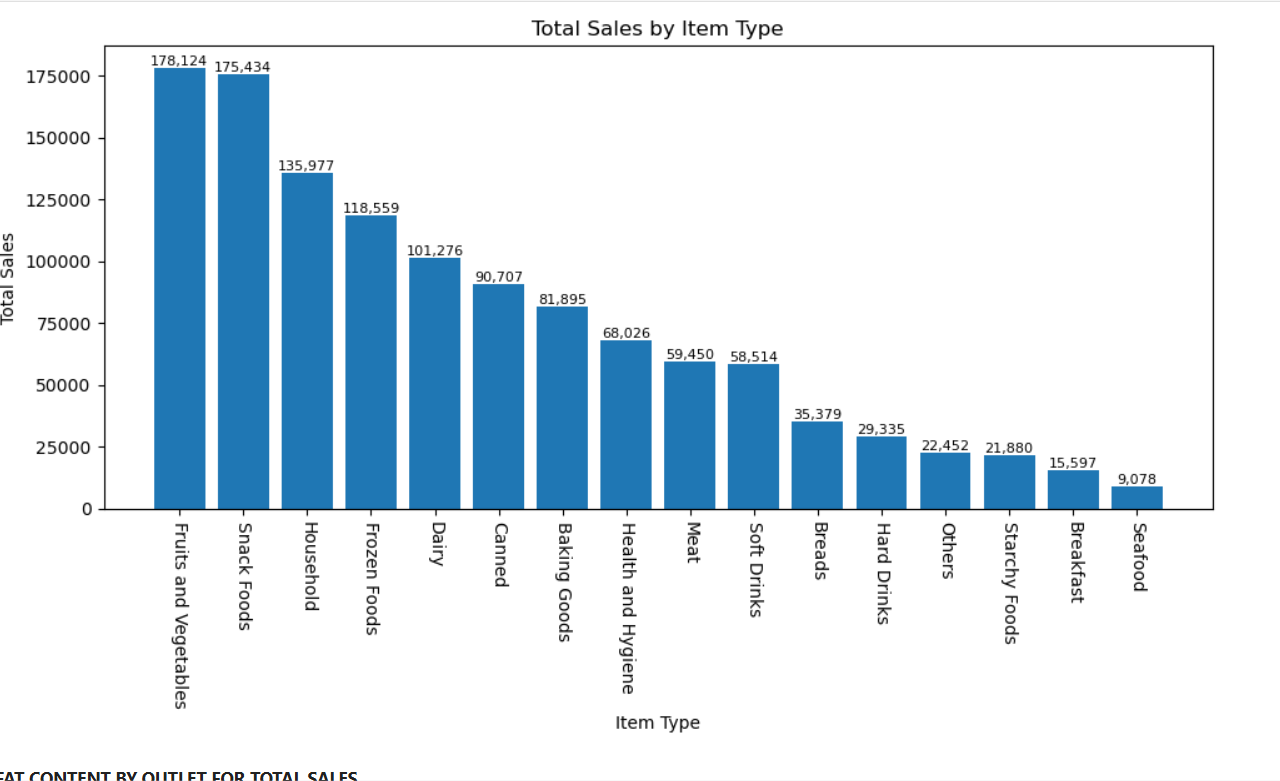
Added annotations and formatted charts for clear communication of sales values.

INSIGHTS

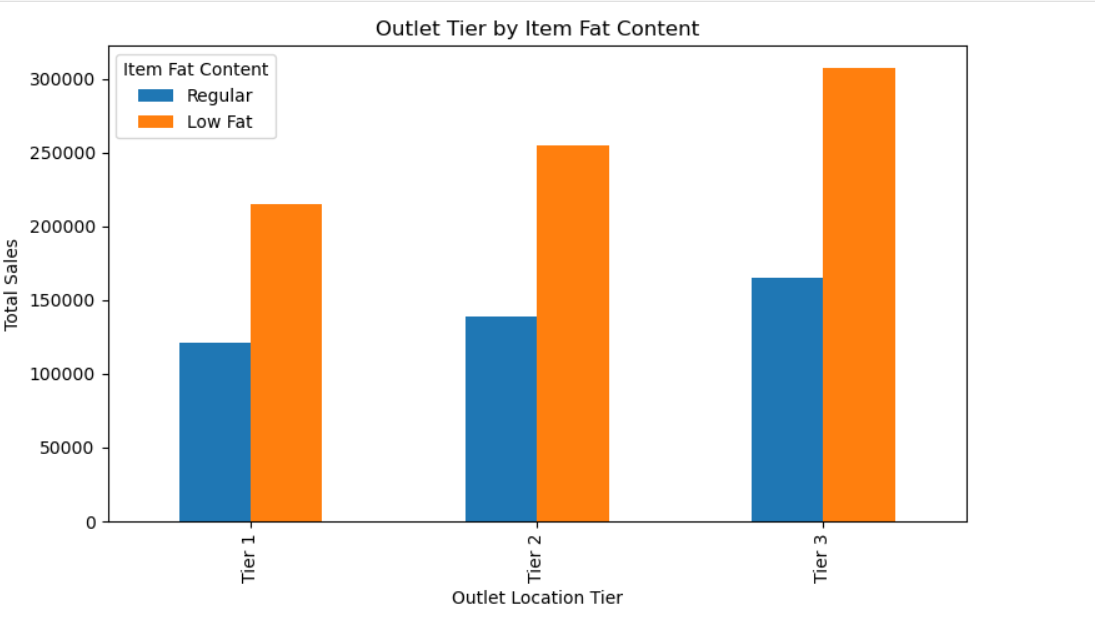
1. **Total Sales by Item Fat Content**  
   This analysis examines the proportion of total sales contributed by products categorized by fat content—namely Regular and Low Fat. The distribution reveals customer preferences towards healthier low-fat options while showing the significant market presence of regular products.



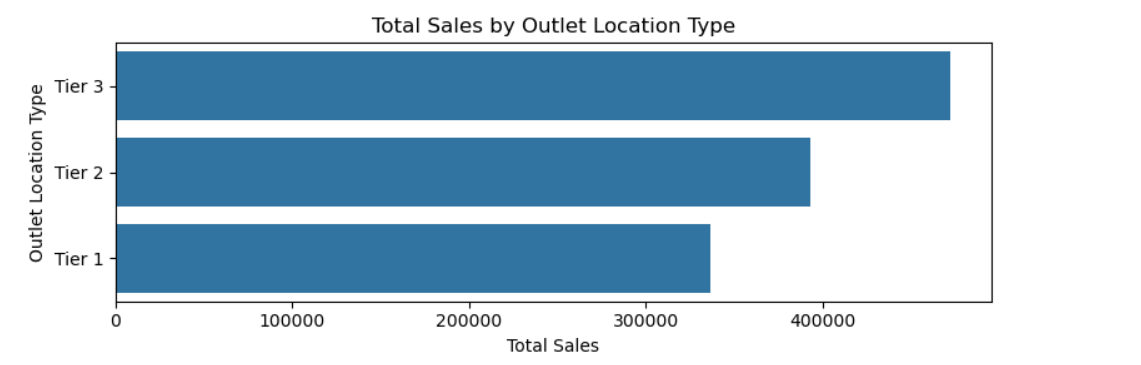
1. **Total Sales by Item Type**  
   Sales are aggregated by product categories to show which types generate the most revenue. Popular categories like Fruits and Vegetables and Snack Foods emerge as key sales drivers, indicating consumer demand and stocking priorities.



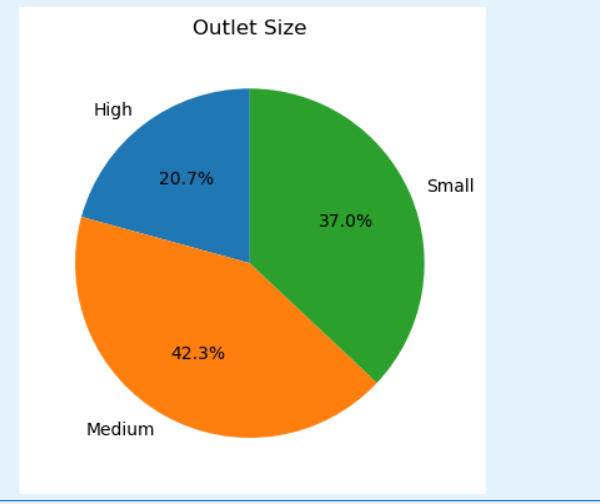
1. **Outlet Tier Sales by Item Fat Content**  
   This section compares sales of Regular and Low Fat products across different outlet tiers. The variation in fat content sales by location tier highlights how customer preferences differ geographically, providing insights for location-focused marketing.



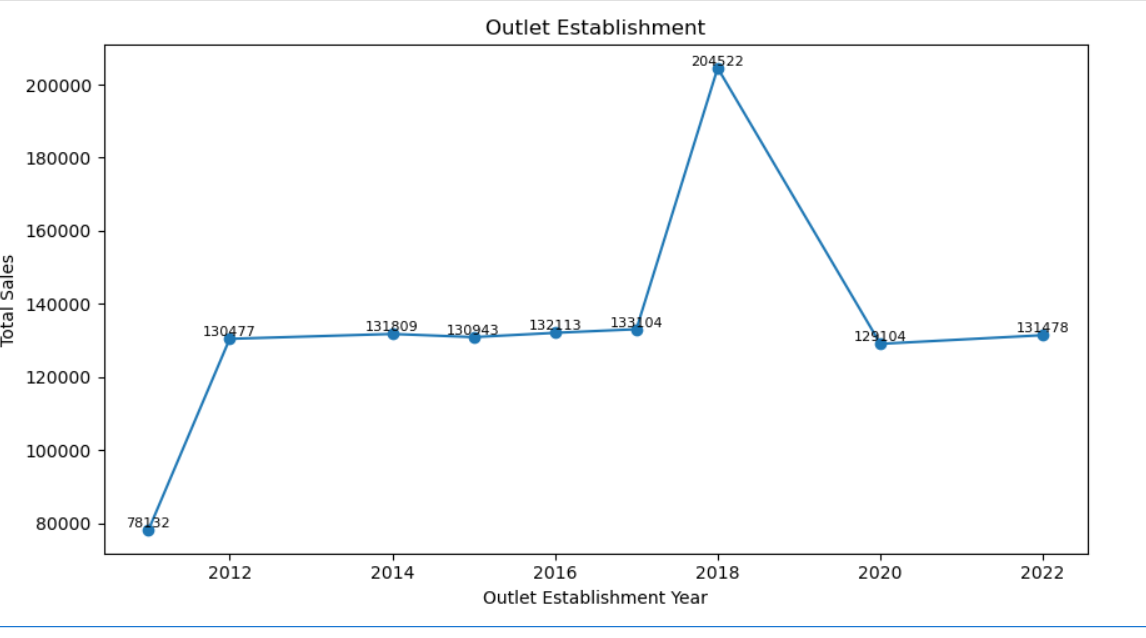
1. **Total Sales by Outlet Location Type**  
   Sales amounts are grouped by outlet location types. The analysis shows that higher tier outlets, usually in urban or suburban areas, dominate revenue generation. This provides guidance for targeted store placement and resource allocation.



1. **Sales Distribution by Outlet Size**  
   Here, sales contributions are analyzed based on outlet size categories. Medium-sized outlets contribute the highest sales share, suggesting an optimal store size for balancing variety and customer reach.



1. **Sales Trend by Outlet Establishment Year**  
   This analysis tracks total sales over outlet establishment years, illustrating how outlet age correlates with sales performance. It can help evaluate store maturity and growth over time.



## 6. Conclusion

The analysis clarifies how product attributes and outlet characteristics impact sales dynamics. Fat content preference and outlet tier location emerge as critical revenue drivers, supported by product popularity and store size factors. The project improved my practical skills in Python data handling and visualization, while delivering actionable business insights.